

UNITED STATES DISTRICT COURT
DISTRICT OF MASSACHUSETTS

UNITED STATES OF AMERICA)	
)	
v.)	CIVIL NO. 07-CV-12056-PBS
)	
JEFFREY SHIELDS)	
)	

DEFENDANT’S POST-DAUBERT HEARING BRIEF

INTRODUCTION

The government proposes to rely on the report and testimony of Dr. Niklos Tomich to meet its burden of demonstrating, by clear and convincing evidence, that Mr. Shields “suffers from a serious mental illness, abnormality, or disorder as a result of which he would have serious difficulty in refraining from sexually violent conduct or child molestation if released.” 18 U.S.C. § 4247(a)(6). Dr. Tomich’s opinion is founded on a review of Mr. Shields’ records and “information from an actuarial measure, the Static 99.” See Niklos Tomich, Report of Forensic Examiner to the Court (“Tomich Rep.”) at 4. Based on this information, Dr. Tomich opines that Mr. Shields “presents with both diagnoses of pedophilia and hebephilia,” and that “Mr. Shields demonstrates serious difficulty in refraining from sexually violent conduct and child molestation and...is likely to engage in future sexual offenses if released to the community at this time.” Tomich Rep. at 13-14.¹

¹ Notwithstanding this forecast, on the witness stand Dr. Tomich claimed that he does not predict future sexual dangerousness: “ I don’t believe I or anyone else can predict the future. We can assume the future, but we can’t predict it.” Tr. at 80. “Well, I don’t predict future sexual dangerousness. I recommend someone as to whether they meet the statutory criteria in the law in sexually dangerous persons cases. I don’t say with any level of certitude that an individual will or will not reoffend; just whether in the Massachusetts case, whether he’s likely to reoffend, and in the federal case, whether he has significant difficulty refraining from reoffending.” Tr. at 162.

Dr. Tomich's report and testimony, and any other expert evidence purporting to show that Mr. Shields suffers a mental abnormality that would make it seriously difficult for him to refrain from molesting children if he is released, should be excluded under Federal Rule of Evidence 702, Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993), and the due process clause of the Fifth Amendment, because the extant methodologies used to predict future sexual dangerousness do not produce results that are sufficiently reliable to be admitted in evidence. Furthermore, even if this Court finds that some method or combination of methods for predicting future sexual dangerousness yield(s) results sufficiently reliable to be admitted in evidence, the methodology that Dr. Tomich *in fact* used in this case does not satisfy the Daubert or due process criteria for admissibility. Nor can his methodology produce results reliable enough to meet the requisite burden of proof – clear and convincing evidence. As defined in O'Malley, Grenig, and Lee, 3 Federal Jury Prac. & Instr. § 104.02 (5th ed):

“Clear and convincing evidence” is evidence that produces in your mind a firm belief or conviction as to the matter at issue. Clear and convincing evidence involves a greater degree of persuasion than is necessary to meet the preponderance of the evidence standard. This standard does not require proof to an absolute certainty, since proof to an absolute certainty is seldom possible in any case.

ARGUMENT

I. EXPERT TESTIMONY ABOUT FUTURE SEXUAL DANGEROUSNESS CANNOT SATISFY THE DAUBERT CRITERIA FOR ADMISSIBILITY.

1. The Evolution of Science and the Law, and the Daubert Criteria for Admissibility.

In Barefoot v. Estelle, 463 U.S. 880, 896-99 (1983), the Supreme Court held that the introduction of expert testimony predicting future dangerousness did not violate due process. See Barefoot, 463 U.S. at 896-99. The Court rejected the view of the American Psychiatric

Association (“APA”) that expert predictions of future dangerousness were almost entirely unreliable.² In declining to bar this testimony, the Court pointedly noted that the rules of evidence in place at the time would generally allow its admission. “[T]he rules of evidence generally extant at the federal and state levels anticipate that relevant, unprivileged evidence should be admitted and its weight left to the fact finder, who would have the benefit of cross examination and contrary evidence by the opposing party.” *Id.* at 898. However, the Court signaled that the evolution of science and the law might cause it to revisit its decision: “We are unconvinced, however, *at least as of now*, that the adversary process cannot be trusted to sort out the reliable from the unreliable evidence and opinion about future dangerousness, particularly when the convicted felon has an opportunity to present his own side of the case.” *Id.* at 901 (emphasis added).

Since Barefoot v. Estelle was decided, the scientific and legal landscape regarding expert testimony on future dangerousness has changed significantly. Over twenty additional years of research now support the APA’s conclusion that expert predictions of future dangerousness are fundamentally unreliable.³ And, in Daubert v. Merrell Dow Pharmaceuticals, the Supreme Court charged federal trial judges with the task of excluding unreliable expert testimony. Daubert v.

² See Amicus Brief, American Psychiatric Association, Barefoot v. Estelle, introduced as Defendant’s Exhibit 4. The APA concluded that “psychiatric predictions of long-term future dangerousness are wrong in at least two out of every three cases.” This amicus brief was based, in part, upon the seminal work of John Monahan cited in the Affidavit of Daniel Kriegman, Ph.D. (“Kriegman Aff”) ¶ 12 & Exhibit 2 (History of Clinical Prediction).

³ See Kriegman Aff., Exhibit 2 (History of Clinical Prediction).

Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).⁴

Under Daubert, the trial court as gatekeeper must ensure that expert testimony is founded on a methodology that produces findings sufficiently reliable to be admitted. Id. at 589. The Court articulated five non-exclusive factors for courts to employ in making the reliability determination: 1) whether the methodology can be and has been tested; 2) whether the methodology has been subject to peer review and publication; 3) the known or potential rate of error in results yielded by the methodology; 4) the existence and maintenance of standards for the implementation of the methodology; and 5) whether the methodology has achieved general acceptance in the relevant scientific community. Id. at 593-94. Under Daubert, the proponent of the expert testimony bears the burden of demonstrating, by a preponderance of the evidence, that the testimony is sufficiently reliable and probative to be admitted. See Daubert, 509 U.S. at 593 n. 10 (citing Bourjaily v. United States, 483 U.S. 171, 175-76 (1987)); Cook ex rel. Estate of Tessier v. Sheriff of Monroe County, Fla., 402 F.3d 1092, 1107 (11th Cir. 2005); Bourne v. Town of Madison, 2007 WL 1447672 at 3 (D.N.H. 2007).

As judged against the Daubert criteria, none of the extant methodologies of predicting future sexual dangerousness produce results sufficiently reliable to be admitted in evidence.

2. The Unreliability of Predictions Based on Actuarial Instruments.

The predictions yielded by even the best actuarial tools are not sufficiently reliable to be admitted in evidence under Daubert, much less to satisfy the clear and convincing evidence

⁴ The Daubert standard replaced the test from Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923). Frye required only that an expert's theory be generally accepted by the relevant scientific community before a court could admit it.

burden of proof of section 4248. Kriegman Aff. at ¶ 31 & Exhibits 7, 10 thereto. Indeed, even the leading researchers devoted to studying actuarial risk assessment instruments express serious reservations about using them to predict future sexual dangerousness for the purpose of civil commitment:

In urging the use of ARA [Actuarial Risk Assessment], we do so against the backdrop of existing SVP laws and serious concerns raised about their constitutionality and wisdom. Although our paper is premised on findings that ARA is superior to clinical assessment methods, we do not take the position that either is sufficient to justify the massive and long-term deprivation of liberty inherent in SVP laws. Our point is simply that *if* courts deprive people of liberty based on assessed risk, *then* ARA should be part of the assessment. Courts should use ARA in part because it will improve risk assessment. But more importantly, from our perspective, ARA brings a transparency that will allow for a clearer understanding of the true nature of risk assessment, *including its significant limits and potential for misuse*.

Eric Janus & Robert Prentky, *Forensic Use of Actuarial Risk Assessment with Sex Offenders: Accuracy, Admissibility and Accountability*, 40 Am. Crim. L. Rev. 1443, 1446 (2003) (emphasis added), attached hereto as Exhibit A.⁵

Thus, as Janus and Prentky indicate, as measured against two of the Daubert factors – peer

⁵ Similarly, the endorsement of actuarial risk assessment by the Association for the Treatment of Sexual Abusers is limited to the proposition that *if* a state chooses to enact civil commitment laws, risk assessment should always be based on the best currently available scientific knowledge. The organization stops far short of vouching for the reliability of actuarial instruments *per se*, and takes no position on the wisdom of civil commitment statutes. See ATSA Rules, introduced as Defendant's Exhibit 1.

review and publication, and acceptance within the relevant scientific community – the status of actuarial instruments is marginal. Although actuarials have been peer reviewed and there are published studies examining their efficacy, the consensus that has emerged from peer review and publication is that their primary contribution to civil commitment proceedings, *if* they are to be used at all, is to highlight the weakness of even the best methods of prognosticating future sexual dangerousness, and *not* to persuade that such prognostications are reliable. Similarly, although actuarials are accepted by the relevant research community as the most reliable of the extant methods of predicting future sexual dangerousness, this agreement as to their *relative* accuracy is *not* an endorsement of the position that they are *per se* “accurate” or “reliable.”

Furthermore, to apply two more of the Daubert criteria, actuarial risk assessments of future sexual dangerousness have been tested, and they must be excluded under Daubert because they have a proven high rate of error.

The use of all actuarial instruments designed to predict future sexual dangerousness begins with a “base rate.” Kriegman Aff. at ¶ 17. A base rate denotes the rate of occurrence of the conduct sought to be predicted within a reference class of individuals similar to the one being studied. Kriegman Aff. at ¶17.⁶ The lower the base rate, the harder it is to be accurate when

⁶ Hanson provides a simple example of the concept of base rate: “The starting point for any risk assessment is the recidivism base rate. The recidivism base rate is the proportion of a group of sexual offenders who will reoffend after a period of time (i.e. the follow-up period). If, for example, 20 out of 100 sexual offenders were reconvicted for a new sexual offense, the recidivism base rate would be 20%. This rate can be used to predict how many offenders will reoffend (e.g. 20 out of 100) as well as to estimate the probability that an individual offender will reoffend (i.e. the ‘typical’ sexual offender has a 20% chance of reoffending).” R. Karl Hanson, et. al., *Sexual Offender Recidivism Risk: What We Know and What We Need to Know* 989 N.Y. Acad. Sci. Ann. 154-55 (2003), attached hereto as Exhibit B.

assessing whether the behavior sought to be predicted – here, future sexual offending – is likely, and the more probable it is that false positives – erroneous predictions of future sexual dangerousness – will be produced. Kriegman Aff. at ¶18; See also Richard Wollert, *Low Base Rates Limit Expert Certainty When Current Actuarials Are Used To Identify Sexually Violent Predators*, 12 Psy. Pub. Pol. & Law 56 (2006), attached hereto as Exhibit C. Thus, if the base rate is low, as it is with sexual offender recidivism, it is especially difficult to accurately predict that a specific sex offender will recidivate, and there is a tendency to overestimate the likelihood that such recidivism⁷ will occur. Thus, using an actuarial instrument to predict future sexual dangerousness, although the most accurate method of making such predictions, is not “accurate” in any meaningful sense of the word:

The development of actuarials changed the way we think about risk prediction. In its increased accuracy over clinical opinion, the actuarials highlight how unreliable clinical judgment is; yet the actuarials produce only limited validity of their own. Actuarials help demonstrate that in the real world of sex offense recidivism prediction, the best predictive method has resulted in only a small degree of validity. Furthermore, this greater validity over clinical opinion highlights how very little correlation there is between an expert’s clinical opinion that someone is sexually dangerous and whether that person will actually recidivate.

Kriegman Aff. at ¶ 30; See also *United States v. Cordoba*, 991 F.Supp. 1199,1204 (C.D.

⁷ “It is more commonly believed that the base rate for sex offender recidivism as a whole lies between 15% and 35%. This base rate fluctuates depending on various factors. For example, studies consistently confirm that older offenders have a significantly lower base rate; on the other hand, offenders under 25 have a higher base rate. Extra-familial child molesters have a higher rate of recidivism than men whose victims were adults. Both of these groups have a higher rate of recidivism than incest offenders, who have the lowest rate of recidivism in this simple classification schema. In all, sex offenders have been shown to have the *lowest* rate of recidivism of all common offense types (only murderers recidivate at a substantially lower rate).” Kriegman Aff. at ¶ 22.

Cal. 1998) (excluding polygraph evidence in part because, “...the court finds the overall error rate is potentially significant, and the error rate for real-life polygraph tests is not known and not particularly capable of analyzing.”).

To illustrate, the Static-99 – an actuarial tool which has been subject to cross-validation studies⁸ – produces inaccurate results when applied to large swaths of offenders. Specifically, the research regarding age and its correlation with reduced recidivism demonstrate that “experts who rely on actuarial tests for predicting likely recidivists for all but the youngest age group will be wrong most of the time.” Wollert, Exhibit C, *supra*, at 71. Using data from Hanson’s 2001 study, Wollert found that the error rate for predicting recidivism using the Static-99 steadily increased as the population being evaluated got older. “[T]his error rate will vary from about 52% for offenders in the 25-29 age range to almost 90% for those in the 60-69 range.” *Id.* He concluded that “current actuarials are of limited value, at best, for [civil commitment] determinations. They may not even be useful at all.” *Id.* Because Mr. Shields is not among the youngest age group, and indeed is among the oldest, using a Static-99 as a basis for predicting recidivism in his case is likely to produce inaccurate results.

Dr. Tomich’s understanding of the Static-99’s accuracy is limited. He first asserted that the predictive accuracy of the Static-99 was “significantly better than chance,” and later said that the Static-99 predicts future sexual dangerousness approximately 25% better than chance:

⁸ A summary of these studies can be found on page 76 of the Revised Static 99 Coding Manual 2003, introduced as Government’s Exhibit 3.

Q: Doctor, the reality is that the best actuarial instrument with the best predictive ability that is known to the science only predicts about 20 percent better than flipping a coin, right?

A: About 25 percent, yes. And, again, it's one data point used in the assessment.

Tr. at pp. 178-179.⁹

⁹In reaching that conclusion, Dr. Tomich demonstrated that he has little or no understanding of the statistical aspects of the research upon which he relies. Thus, whether or not Dr. Tomich's ultimate conclusion that the Static-99 generally predicts about 20 percent better than chance is correct, his method of arriving at that determination is wholly unsound. Dr. Tomich testified as follows:

THE WITNESS: Okay, okay. Uhm. . . the Static-99 has been cross-validated on numerous occasions. And in the back of the coding manual on one of the pages, I believe Page [76] we looked at yesterday, the coefficients that indicate validity are provided. Those are called D statistics. It's roughly twice the size of a correlation. It's used because it gets rid of the base rate, which is (Inaudible) unknown. The base rate is defined as how many sex offenses occur in the community. So Hanson and others utilize the D statistic to do that. D statistics could range from .0 to 1. .5 is the midrange. What we find in the Hanson cross-validation studies and others' cross-validation studies is, depending on the samples used, the D statistics in the area under the curve range from .6 to .9 with an average of .71.

THE COURT: Does that in your mind mean that it's reliable?

THE WITNESS: Yes, it does.

THE COURT: And in English, what does that mean, a D statistic of .71?

THE WITNESS: It means that you're getting results of predictive accuracy, significantly better than chance.

THE COURT: Which is .5?

THE WITNESS: Which is .5.

Q: Doctor, you can put this in laymen's term. You can tell us exactly how much better than flipping a coin using the actuarial instrument is.

A. I think I just did.

Q. Would you do that in laymen's terms. How much more predictive –

A. About 25 percent better.

Tr. 177-78.

Dr. Tomich's understanding of the "d" statistic is plainly wrong. "d", otherwise known as "cohen's d" is the standardized mean difference. "In other words, d measures the average

Even assuming that Dr. Tomich was right that the Static-99 generally predicts 25% better than chance, he failed to take into account the impact of base rate on the instrument's accuracy. Dr. Tomich never even attempted to translate Mr. Shield's Static-99 score of 6 into a base rate for the group most like Mr. Shields, nor did he attempt to adjust that base rate for age to accurately ground his assessment. Had he done so, he would have been confronted with two realities: 1) that the appropriate base rate in this case was generally low; and 2) that the unadjusted Static-99 would likely not predict accurately given Mr. Shields' age. His statement that the Static-99 predicts 25% better than chance is thus demonstrably false in Mr. Shields' case.

difference between the recidivists and the non-recidivists, and compares this difference to how much recidivists differ from each other, and how much non-recidivists differ from each other." Hanson, *Meta-Analysis 2007-01*, Defendant's Exhibit 5 *supra*, at 7; See also R. Karl Hanson & Kelley E. Morton-Bourgon, The Characteristics of Persistent Sexual Offenders: A Meta-Analysis of Recidivism Studies, 73 J. Consulting & Clin. Psy. (2005) at 1156, attached hereto as Exhibit D. "The d statistic . . . is less influenced by recidivism base rates than correlation coefficients—the other statistic commonly used in meta-analyses." *Id.* "According to Cohen (1988), d values of .20 are considered "small", .50 "medium", and .80 "large". *Id.* Because a "d" value of .00 represents pure chance, the higher the d value of a given variable, the stronger relationship with recidivism that variable is demonstrated to have.

Dr. Tomich testified that a d value of .5 is pure chance. That is not the case; a d value of .00 is pure chance. He also testified that having a d value of .70 meant the factor was reliable. That was wrong; a value of .70 simply means that the variable has a medium or moderate relationship with recidivism. Finally, he testified that he could calculate how much better than chance a given variable was by simply knowing its d value. That was also erroneous. The assertion is nonsensical because the d statistic—or any other statistic for that matter—cannot be used or manipulated in such a way. Indeed, in earlier testimony, Dr. Tomich admitted that he simply does not understand the d statistic. When asked by the Court to explain the concept of statistical significance as it relates to sex offender risk assessment using the plethysmograph, Dr. Tomich was at a loss, saying "I don't know why they're [D statistics] statistically significant. I don't think anybody does. It's just been found that they are." Tr. at 55.

3. The Adjusted Actuarial Approach: Even Less Reliable Than Actuarials Alone.

Since the development of actuarial instruments, various approaches of combining their use with the use of clinical judgment have evolved. The two main approaches are called the “guided clinical method” and the “adjusted actuarial method”. Kriegman Aff. at ¶ 32.

In the guided clinical method (sometimes called structured clinical or structured professional), evaluators consider a wide range of empirically validated risk factors and then form an overall opinion concerning the offender’s recidivism risk. In this approach, the method for translating the identified risk factors into recidivism rates is not explicitly determined. See R. Karl Hanson, *What Do We Know About Sex Offender Risk Assessment?*, 4 Psy. Pub. Pol & Law 50, 52-53 (1998), attached hereto as Exhibit E. In the adjusted actuarial method the evaluator begins with an actuarial prediction but may then adjust that prediction after considering empirically validated factors that were not included in the actuarial measure. Id. at 53. See also Kriegman Aff. at ¶ 32.

As Dr. Kriegman has detailed in his affidavit, the research indicates that neither of these methods are as accurate as the pure actuarial method, which itself provides limited accuracy. Kriegman Aff. at ¶ 32 & Exhibit 8 thereto (Clinical Variations). The most recent research indicates that the accuracy of evaluations based on structured professional judgment was intermediate between the accuracy of pure actuarial assessments and the accuracy of unstructured professional judgments, which are consistently the least accurate method of risk assessment. Kriegman Aff. & Exhibit 8 thereto (Clinical Variations); R.

Karl Hanson, *The Accuracy of Recidivism Risk Assessments for Sexual Offenders: A Meta-Analysis* 2007-01 (2007) at I, introduced as Defendant's Exhibit 5. Thus, while adjusting actuarial assessments may at times be necessary and can be defended in certain circumstances, there is no empirical evidence that this improves predictive accuracy and considerable evidence that the opposite is true. See Kriegman Aff. Exhibit 8 (Clinical Variations) (citing studies).

4. Courts and Commentators are Increasingly Skeptical of Future Dangerousness Predictions.

The Daubert decision and the accumulating research showing the unreliability of predictions of future dangerousness have led some courts to conclude that expert testimony regarding future dangerousness is now, contrary to Barefoot, inadmissible under the Federal Rules of Evidence. For example, writing in a capital case in which the government alleged the aggravating factor of future dangerousness, the District Court of Massachusetts (Wolf, J.) engaged in a lengthy analysis of the intersection between Daubert and expert testimony regarding future dangerousness, concluding that “[d]evelopments in the law and more recent scientific research suggest that expert testimony on future dangerousness would be *inadmissible* under the Federal Rules of Evidence . . .” United States v. Sampson, 335 F.Supp.2d 166, 218 (D. Mass. 2004) (emphasis added).¹⁰ See also Flores v. Johnson, 210 F.3d 456, 464 (5th Cir. 2000) (Garza, J.,

¹⁰In Sampson the issue before the Court was whether the aggravating factor of future dangerousness should be submitted to the jury. The Court acknowledged that its decision to submit this aggravating factor to the jury was controlled by the United States Supreme Court's decision in Jurek v. Texas, 428 U.S. 262 (1976), which upheld a death penalty statute with an aggravating factor of “future dangerousness.” However, the Court opined that “[t]his court's experience in this case, however, suggests that it may now be appropriate for the Supreme Court to revisit Jurek.” Sampson, 335 F.Supp.2d at 218. The Court went on to discuss the reliability of expert testimony regarding future dangerousness in great detail, concluding “[t]hus, this court

concurring) (“On the basis of any evidence thus far presented to a court, it appears that the use of psychiatric evidence to predict a murderer’s ‘future dangerousness’ fails all five Daubert factors.”). In addition to these cases, a variety of law review articles have questioned the continuing admissibility of future dangerousness predictions after Daubert.¹¹

The cases cited by the government, which purportedly show that actuarial predictions of future dangerousness remain admissible after Daubert, do nothing of the sort and are not contrary to the trend, noted *supra*, toward the exclusion from evidence of future dangerousness predictions under the Daubert framework. See Government’s Opposition to Renewed Motion for Daubert Hearing, at pg. 5 (“Opposition”).

In fact, *none* of the cases cited by the government in its Opposition actually analyzes actuarial risk assessments using the Daubert factors. See In re Detention of Thorell, 149 Wash.2d 724, 72 P.3d 708 (2003), cert denied, 541 U.S. 990 (2004) (court was not required to conduct Frye hearing on the admissibility of actuarial instruments in a sexually dangerous person proceeding); In re Commitment of Simons, 213 Ill.2d 523, 535-36, 821 N.E. 2d 1184 (2004)

would probably have excluded any expert evidence offered on future dangerousness because its probative value would have been outweighed by the danger of creating unfair prejudice.” Id. at 220.

¹¹See Clayton Skaggs, *Note, Kansas’ Sexual Predator Act and the Impact of Expert Predictions: Psyched Out by the Daubert Test*, 34 Washburn L.J. 320, 342 (1995) (assessing psychiatric predictions under the four identified Daubert factors, and concluding “a trial judge applying the Daubert Court’s ‘general observations’ would find the expert predictions inadmissible”); Erica Beecher-Monas & Edgar Garcia-Rill, *Danger at the Edge of Chaos: Predicting Violent Behavior in a Post-Daubert World*, 24 Cardozo L. Rev. 1845 (2003); Thomas Regnier, *Barefoot in Quicksand: The Future of “Future Dangerousness” Predictions in Death Penalty Sentencing in the World of Daubert and Kumho*, 37 Akron L. Rev. 469, 488 (2004). See also Michael H. Gottesman, *From Barefoot to Daubert to Joiner: Triple Play or Double Error*, 40 Ariz. L. Rev. 753, 755 (1998) (arguing generally that Daubert will require the exclusion of psychiatric predictions and that Daubert cannot be squared with Barefoot.).

(actuarial risk assessments satisfy the Frye standard); In re Commitment of R.S., 173 N.J. 134, 136, 801 A.2d 219, 221 (2002) (“We resort here to the Frye standard because of the liberty interest at stake.”). Indeed, the only case cited by the Government that even addresses Daubert, and then only briefly, is In re Detention of Holtz, 653 N.W.2d 613, 619 (Iowa App. 2002) (en banc).

In Holtz, after pointedly noting that the Iowa rules of evidence and case law made it clear that “we are committed to a liberal view on the admissibility of expert testimony, and we have been quite deferential to the district court in the exercise of its discretion in that area,” the court went on to review the principles governing the admissibility of expert testimony in Iowa. Holtz, 653 N.W.2d at 615 (internal citation omitted). In so doing, the Holtz Court noted in passing that “trial courts are not required to apply” the Daubert analysis in assessing the admissibility of expert testimony, although they might find it useful to consider one or more of the Daubert factors “if deemed helpful in a particular case.” Id. at 615-16.

In the case before it, the Holtz court did *not* found its decision on an application of the Daubert factors. Instead, it rested its decision to affirm the trial court’s admission of expert testimony about actuarial risk assessment on two grounds: 1) a review of precedents from other state courts; and 2) the fact that “the district court ably summed up the issues...” Id. at 619. It is therefore worth examining these parts of the Holtz case more closely.

As to the state court authorities, the Holtz Court did not note whether any of them employed a Daubert analysis in admitting expert testimony regarding actuarial instruments but, if any of them did, this clearly did not influence the Holtz Court’s decision to follow them. See Id.

at 619 & n. 5. It evidently was not the mode or quality of analysis in these decisions that impressed the Holtz Court, but simply their number. See Id. at 619 (agreeing with a New Jersey decision that “[o]ur research has revealed no state appellate court decision which has found actuarial instruments inadmissible at SVP proceedings.” (internal citation omitted)).

Turning to the trial court’s statement of the issues surrounding admissibility, which the Iowa Court of Appeals endorsed as its second ground for affirmance, the following representative sections are noteworthy:

...he [the expert] does not represent them [actuarial instruments] to be an end-all... he relies on them to corroborate as a check and balance, a starting point... all of the expert witnesses in that field who testify, testify similarly. They all indicate that those tests can be used, that they are used... I agree that it’s very important that once those measuring tools are discussed and introduced into a case that full and complete cross-examination and disclosure about the limitations of those tests be disclosed to the jury...The tests do contain many factors that experts in the field rely on in order to make risk assessments as to recidivism.

Id. at 618-19 (no internal citation provided).

It is thus clear that the trial court admitted the expert testimony about actuarial instruments for three reasons: 1) the testifying experts used them in conjunction with other techniques of predicting future sexual dangerousness; 2) the trial court intended to permit thorough cross-examination about the actuarials; and, clearly most importantly, 3) the trial court found that the use of actuarials was accepted by experts in the field of sexual offense risk prediction. Far from a Daubert-type analysis, this is essentially a slightly modified application of Frye’s general acceptance test.

The Government's contention that the Holtz case, or any of its other cited authorities, stands for the proposition that states continue to admit expert testimony about actuarial risk assessment under Daubert thus simply cannot withstand scrutiny.

And, the Government's cited authorities are in keeping with the state of the case law generally. No state has actually *analyzed* expert testimony predicting future sexual dangerousness generally, or the use of actuarial instruments to make such predictions specifically, using the Daubert criteria. Indeed, as one commentator noted, "many (but not all) Frye-jurisdiction courts have simply admitted clinical dangerousness testimony without any sort of vetting to insure its scientific bona fides." See Janus & Prentky, Exhibit A *supra*, at 1482-83 (describing different approaches state courts have taken in analyzing future dangerousness testimony).^{12 13}

¹² Additionally, state courts evaluating actuarial risk assessments have often demonstrated a fundamental misunderstanding of the science involved. For example, in In re Commitment of R.S., the reviewing court distinguished between cases where an evaluator relied solely upon clinical judgment to predict future dangerousness and cases where an evaluator used actuarial instruments. The court held that if an evaluator relied on actuarial instruments, the State was required to satisfy the Frye standard, but if the evaluator relied upon clinical judgment alone, the expert was permitted to testify as to his or her opinions without satisfying Frye. In re Commitment of R.S., 339 N.J. Super at 537-538; Holtz, 653 N.W.2d at 619. This approach is "incoherent, because it allows for the possibility that the more reliable ARA [actuarial risk assessment] would be treated more strictly than the less reliable CRA [clinical risk assessment]" and thus demonstrates a fundamental lack of understanding of the scientific findings regarding the validity and accuracy of various risk assessment methods. Janus & Prentky, *supra*, at 1483.

¹³ A review of the states with civil commitment laws reveals that some states subject actuarial risk assessments to Frye's general acceptance analysis, while other states *exempt* such testimony from the Frye analysis entirely, and still other states, like Massachusetts, simply legislate the admissibility of such testimony, thereby bypassing any judicial scrutiny of such evidence. "These provisions are a very radical departure from the rules of evidence that ordinary govern, but the Legislature has made its policy clear in this regard, and it is not for us to quarrel with it." Commonwealth v. Bradway, 62 Mass. App. Ct. 280 (2004) (denying a Daubert-Lanigan challenge to the testimony of a qualified examiner because qualified examiners are presumed competent to testify by statute). It is worth noting that, while Dr. Tomich is a qualified examiner, having been so appointed by the Massachusetts Department of Corrections, and is thus presumed

II. DR. TOMICH IS NOT QUALIFIED TO TESTIFY AS AN EXPERT AND HIS METHODOLOGY PRODUCES PREDICTIONS OF FUTURE SEXUAL DANGEROUSNESS THAT ARE INSUFFICIENTLY RELIABLE TO BE ADMITTED.

1. Dr. Tomich Lacks the Qualifications to Testify as an Expert in Sex Offender Risk Assessment.

Federal Rule of Evidence 702 requires that an expert be qualified by “knowledge, skill, experience, training, or education” to assist the trier of fact in understanding evidence or determining a fact in issue. Fed. R. Evid. R. 702. To pass muster, the proposed expert’s qualifications must be “in the specific subject for which his testimony is offered.” Sutera v. Perrier Group of America, Inc., 986 F.Supp. 655, 661 (D.Mass. 1997); See also Berry v. City of Detroit, 25 F.3d 1342, 1351 (6th Cir. 1994) (“The issue with regard to expert testimony is not the qualifications of a witness in the abstract, but whether those qualifications provide a foundation for a witness to answer a specific question.”).

The proponent of the expert testimony bears the burden of demonstrating that the proffered expert is qualified to deliver it. See Cook ex rel. Estate of Tessier v. Sheriff of Monroe County, Fla., 402 F.3d 1092, 1107 (11th Cir. 2005); Bourne v. Town of Madison, 2007 WL 1447672 at 3 (D.N.H. 2007). The expert’s qualifications must be established by a preponderance of the

competent by statute to testify as an expert in Massachusetts state courts under M.G.L. c. 123A, he has never been qualified as an expert in Federal court, nor are his qualifications subject to the scrutiny of the Massachusetts courts because of the legislated presumption of competence of “qualified examiners” under Massachusetts law. Tr. at 35.

evidence. See Daubert, 509 U.S. at 593 n. 10 (citing Bourjaily v. United States, 483 U.S. 171, 175-76 (1987)).

Here, the government proposes to present Dr. Niklos Tomich, to testify that, as an expert in the diagnosis of sex offenders, and in the assessment of their risk of recidivism, he has concluded that Mr. Shields meets the definition of a sexually dangerous person. As even a selective review of Dr. Tomich's curriculum vita and testimony demonstrates, the government cannot meet its burden of proving that Dr. Tomich is qualified to opine on these subjects.

First, many of Dr. Tomich's credentials may be of questionable value. His undergraduate degree is from a Y.M.C.A. college that no longer exists, and the curriculum he studied may or may not have been accredited. See Tr. at 119; Exhibit 1 at 1. He took his graduate degree from the Chicago School of Professional Psychology, an institution that apparently did not require him to take the Graduate Record Examination as a prerequisite to admission. Tr. at 126. Once there, he undertook a doctoral program that did not receive even provisional accreditation until 1988, the year he was awarded his degree. Exhibit 1 at 1.; <http://www.theChicagoSchool.edu/content.cfm/history>.

Second, Dr. Tomich has never performed any research in the sex offender field, has not written a single peer-reviewed article, and has never served on a peer-review board. Tr. at 130. Rather, he has spent his entire educational and professional life as a clinician and administrator. His principal academic credential, a Psy.D., is, by his own description, "a practice degree" that he took because he was "not interested in conducting research." Tr. at 125; Tomich C.V. entered as Government's Exhibit 1 at 1 (and hereinafter "Exhibit 1"). Dr. Tomich's professional experience

reflects his disinterest in research and consists entirely of positions in which he has provided treatment to or evaluations of the mentally ill, or supervised and organized the activities of those who do the same. See Exhibit 1 at 1-4. Yet, the government seeks to present him as an expert able to provide the finder of fact with guidance as to the merits of various methodologies used to predict future sexual dangerousness and the weight that should be afforded to those methodologies in order to support his “expert” opinion that Jeffrey Shields is a sexually dangerous person. This Court should not permit that testimony.

Finally, the sole criteria for the “qualified examiner” credential that Dr. Tomich enjoys in Massachusetts is that one have at least two years of experience treating or evaluating sex offenders. Tr. at 110-11. The appointment itself is made, not by a mental health professional capable of fully evaluating the prospective examiner, but by the Commissioner of Corrections. Id.

These examples reveal serious shortcomings in Dr. Tomich’s education, training, and experience that undercut his proffered status as an expert in the evaluation of sex offenders and the assessment of their risk of recidivism. To testify on those subjects, Dr. Tomich would have to have a much better understanding of the research regarding sex offender risk assessment than he does, and an understanding of the methodologies underlying that research. That he does not possess those qualifications is apparent from his own testimony.

To take one example, when asked by the Court to explain the concept of statistical significance as it relates to sex offender risk assessment using the plethysmograph, Dr. Tomich was at a loss, saying “I don’t know why they’re [D statistics] statistically significant. I don’t think anybody does. It’s just been found that they are.” Tr. at 55. At another point in his testimony,

Tomich seemed aware of the limitations of his ability to understand and effectively communicate some of the science of risk analysis:

Q: And if you are not a statistician, do you believe that you are competent to offer testimony with respect to statistical analysis?

A: To the degree that I can. I mean, if I –

Q: Well, that's what I'm trying to find out. What is the degree that you can, Doctor?

A: If I'm asked a question that I don't understand or feel I can't answer, I think I can say that.

Tr. at 130-31.

Simply put, Dr. Tomich's education and professional experience are those of an average clinician. He knows the techniques that he is supposed to employ to achieve a desired result, but he does not understand, and consequently cannot articulate, why those techniques produce the result – or why they might not. See General Electric Co. v. Joiner, 522 U.S. 136, 146 (1997) (“[N]othing...requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert.”). In similar situations, courts have found proffered experts not qualified. Cf. Bogosian v. Mercedes-Benz of North America, Inc., 104 F.3d 472, 476-78 (1st Cir. 1997) (master mechanic not qualified to opine on alleged design defect in automobile transmission parking mechanism); Sutera, 986 F.Supp. at 687 (oncologist and hematologist not qualified to testify to genesis of plaintiff's leukemia because no expertise in epidemiology, toxicology, biostatistics or risk-assessment).

The Court should do the same here, ruling that Dr. Tomich is not a qualified expert. However, even if the Court concludes that he is qualified, the following examples show that Dr. Tomich's testimony must nevertheless be excluded because the methodology underlying it is fundamentally unsound.

2. Dr. Tomich Misdiagnosed Mr. Shields with "Hebephilia," a Rejected Diagnostic Category.

Dr. Tomich diagnosed Mr. Shields with both pedophilia and hebephilia. Tomich Rep. at 13. Although Dr. Tomich admitted that hebephilia is not a diagnosis listed in the DSM-IV, he asserted that "it is nevertheless an often used diagnostic term that is indicative of a deviant pattern of sexual arousal to adolescent individuals under the age of consent." *Id.* Dr. Tomich opined that the hebephilic arousal becomes "deviant" when the subject is under the legal age of consent and the arousal is acted upon. Tr. at 93.

Drs. Plaud and Kriegman disagreed with Dr. Tomich's contention that the term "hebephilia" denotes a mental abnormality. Dr. Plaud noted that the terms hebephilia and ephebophilia have been "purposefully rejected as appropriate psychodiagnostic labels within the framework of sexual disorders." Report of Joseph J. Plaud ("Plaud Rep.") at 13. Dr. Kriegman similarly noted that the American Psychiatric Association rejects hebephilia as a diagnostic category. Tr. at 252. Both doctors maintain that an attraction to adolescents is a normal pattern of arousal rather than a deviant pattern of arousal. "There is no underlying sexual arousal disorder here; rather, the issue is one of statutory issues related to age of consent for sexual behavior." Plaud Rep. at 13. See also Tr. at 252.

“The DSM-IV-TR is almost universally relied on as the authoritative support for expert opinions on mental abnormality or personality disorder. The classification of a syndrome as a mental disorder in the DSM-IV-TR must be regarded as the primary standard for medical validity in the SVP context.” Robert Prentky & Eric Janus, et. al., *Sexually Violent Predators In The Courtroom: Science on Trial*, 12 Psy. Pub. Pol & Law, 357, 364 (2006), attached hereto as Exhibit F. Dr. Tomich acknowledged that the DSM-IV is generally accepted as authoritative within the community of psychologists and psychiatrists. Tr. at 88.

A diagnosis of a serious mental illness, abnormality, or disorder is a statutory and constitutional prerequisite to commitment as a sexually dangerous person. 18 U.S.C. §4248; See also Kansas v. Hendricks, 531 U.S. 346, 369 (1997) (statutory requirement of “mental abnormality or personality disorder” as predicate for commitment satisfies substantive due process requirements). In SDP proceedings, diagnosis is directly tied to indeterminate deprivation of liberty. Because the legitimacy of SDP proceedings hinges on the scientific integrity of the diagnostic decision, Dr. Tomich’s willingness to use a diagnostic term that does not connote deviant arousal and was purposely not included within the DSM-IV is troubling, as it reflects his willingness to use principles that are considered to be scientifically unsound:

When an appropriate clinically defensible diagnosis is not available, defendants are occasionally shoehorned into a legitimate diagnosis by ignoring elements required for classification . . . or worse, placing defendants into newly created categories, such as paraphilia NOS - nonconsent, that have no known empirical support. Force-fitting a diagnosis or creating a new DSM diagnosis to justify commitment is clearly unethical for psychologists (Ethical Standards, 9.01; American Psychological Association, 2002, p. 1071). . . when new categories that have never been subjected to empirical scrutiny are relied on, all of the precepts of good science are violated, from the most elementary (can cases be assigned to this

new category reliably) to the more recondite (what does assignment to this category inform us about reoffense risk).

See Prentky & Janus, Exhibit F, *supra*, at 369-70 (2006). That the diagnosis of hebephilia is scientifically unsound is illustrated by the fact that evaluators using this diagnosis cannot define when the normal hebephilic arousal becomes “deviant”. Evaluators suggest that the answer may lie in the degree to which the individual is impaired by that arousal, and, like Dr. Tomich, look to whether the individual has continued to act on his attraction despite previous prosecution, or the fear of prosecution, to determine if he is impaired. This analysis overlooks the fact that the age of consent varies between jurisdictions. Thus, what is evidence of “deviant” sexual arousal in one jurisdiction may not be evidence of “deviant” sexual arousal in another jurisdiction. Hence, the term “hebephilia” can never be consistently defined and, consequently, it cannot be empirically studied. Accordingly, it can never provide a valid basis for civil commitment.

3. Dr. Tomich Did Not Use an Adjusted Actuarial Approach.

Dr. Tomich testified that he uses the “adjusted actuarial approach,” and that under this approach he utilizes the Static-99 as one “data point” in his analysis. Tr. at pp. 58, 133. “It grounds the opinion in risk factors that have been researched and empirically derived that are indicative of risk, and then that’s adjusted based on dynamic risk factors and protective factors that might offset risk as well as historical factors.” Tr. at pg. 133.

As a number of examples will illustrate, Dr. Tomich did not *in fact* use an adjusted actuarial approach, at least not as that methodology is understood in the literature. In particular, Dr. Tomich deviated from a true adjusted actuarial approach by adjusting his assessment of Mr. Shields’ future sexual dangerousness using factors that research has not shown to bear any

relationship to sexual dangerousness or that research has positively demonstrated *do not* impact the likelihood of sex offender recidivism. Additionally, Dr. Tomich used some empirically validated factors in ways that are not approved in the literature and are incompatible with a true adjusted actuarial approach.

A. Misuse of the Age Factor

Although Dr. Tomich purported to use the “adjusted actuarial method,” he declined to actually *adjust* Mr. Shields’s recidivism base rate using the only dynamic factor that has received the most empirical validation - age. Although Tomich averred that he recognized Hanson as a leader in the field on recidivism and age, Tr. at 48; 188, he did not follow the methodology for age-related adjustment of the Static-99 detailed by Hanson. Such an adjustment would have reflected a reduction in Mr. Shields’ risk to reoffend based on his membership in the 40-49 year old age group. Instead, Dr. Tomich first identified Mr. Shields’s age - 46 - as a potential “protective” factor. He then proceeded to wholly discount this factor:

In addition to risk factors, this evaluator also considered protective factors that Mr. Shields may present with at the current time. Protective factors are those factors which may diminish or ameliorate risk. Mr. Shields is now 46 years of age. Research has demonstrated that as individuals age, their risk decreases. However, child molesters are known to reoffend well into their fifties. Additionally, in Mr. Shields’ particular case, he was 27 at the time of his first set of sexual offenses, 37 at the time of his sexual offense in 1998 and 41 at the time of the index sexual offense. The data clearly indicated that age in and of itself would not be a protective factor in Mr. Shields’ case.

Tomich Rep. at 15.

In 2006, Hanson published a peer-reviewed article entitled “Does Static-99 Predict Recidivism Among Older Sexual Offenders?” This study showed that the Static-99 did not

sufficiently capture declines in recidivism risk associated with age, and concluded that older offenders had lower sexual recidivism rates than would be expected based on their Static-99 scores. See R. Hanson, *Does Static-99 Predict Recidivism Among Older Sexual Offenders?* 18 Sex Abuse: A Journal of Research & Treatment, 343 (2006), attached hereto as Exhibit G.¹⁴ This study provides an evaluator with empirically validated tools to adjust the recidivism base rate based on the age of the offender at the time of his release from custody on his “index offense,” or the offense triggering his most recent incarceration.¹⁵

To illustrate, under the 2003 Static-99 coding rules, a score of 6 indicates that the individual scored similar to a test group which had a recidivism rate of 39% at 5 years post-release, 45% at 10 years post-release, and 52% at 15 years post-release.¹⁶ However, Hanson’s 2006 study demonstrates that, for men like Mr. Shields who are between the ages of 40-49 at release from custody, the age-adjusted recidivism base rate at 5 years post-release is actually 25.7% – not 39% as reported in the original Static-99. *Id.* at 350. Thus, the original Static-99 *overestimates* Mr. Shields’ risk of reoffending.

¹⁴ Sex Abuse: A Journal of Research & Treatment is published by the Association for the Treatment of Sexual Abusers (ATSA).

¹⁵ This study is a follow up to Hanson’s 2002 study of age and recidivism. See R Karl Hanson, *Recidivism and Age: Follow-Up Data From 4,673 Sexual Offenders*, 17 Jour. Interpersonal Violence (2002), introduced as Defendant’s Exhibit 6. The 2002 study provided raw data that evaluators could develop to adjust an older offender’s recidivism base rate. The example provided by Dr. Kriegman in his testimony quoted below uses “10 year” data from Hanson’s 2002 study. In 2006 Hanson took the available data and published revised base rate estimates for the Static-99 for older offenders for the period covering 5 years post-release.

¹⁶ Government’s Exhibit 3, at 69.

Hanson's 2006 study gives evaluators a simple chart to adjust base rate to account for offender age at 5 years post-release. It is worth noting that there are other, valid methods to adjust the Static-99 to account for an offender's age.¹⁷ In his testimony before the Court, Dr. Kriegman detailed an alternative way of using age to adjust the Static-99 using 10 year post-release recidivism figures, and emphasized the virtues of applying the age-adjusted scoring method correctly and rigorously:

Well, after giving the fact-finder the caveat of the limited validity of my opinion and our tools, I would start with the Static-99. And I would not use the 15-year figures because in 15 years he'll be 62, which is down into the period where its almost impossible to find a recidivist. Certainly, if he gets through probation, he'll be 49. In any case, I think the 10-year figures are the most you could possibly use, and that starts us at a recidivism risk of 45 percent. Then, using the aging data for extrafamilial child molesters, we see that on the chart there an age reduction from the age of 35 to age of 47 or 49 at the end of probation, we see over a one-third reduction in risk on the rate of recidivism. So I would take one-third off the 45, a little bit more than that. So you would have a best prediction of under 30 percent for men who scored a 6 on the Static-99 and are at Mr. Shields' age and are extrafamilial child molesters. And there's no clinical judgment in that. It's all transparent. You can see exactly how I'm weighing it and why, what the empirical data is."

Tr. at 241.

These are all valid ways of making adjustments to arrive at an accurate base rate. Dr. Tomich, however, simply declined to use *any* method available to adjust the Static-99 base rate, ignored the age-related data completely, and discounted age as a protective factor entirely. Dr. Tomich has stressed that the Static-99 is a key "data point" in his analysis. Thus, by relying on an unadjusted Static-99 base rate that over-reports Mr. Shields' risk of reoffending he relies on a

¹⁷ For example, Prentky & Janus suggested a method of adjusting base rates which would indicate a reduction in recidivism risk of approximately 2% per year after age 40. Prentky & Janus, *Sexually Violent Predators In The Courtroom*, Exhibit F, *supra*, at 374.

“data-point” that is based on inaccurate information, and thereby prejudices Mr. Shields by using a methodology that is without scientific justification.

Dr. Tomich’s analysis regarding age suffers from two additional problems. First, he suggested that age could be wholly discounted was because Mr. Shields had, in his view, “recently” reoffended. He said, “I take that scientific information and weigh it against the facts and circumstances of the specific case at hand. So an individual may be 55 years old, but if his last offense was when he was 52 years old, it certainly says something about his reoffending at age 55.” Tr. at 49.

Tomich ignores the fact that the correct method of adjusting the base rate of the older offender is based on the offender’s age *at release*, as explained, *supra*. To accurately adjust for age in an empirically validated fashion, the evaluator simply looks at the offender’s score on the Static-99 and his age at release. There is no empirically validated coding rule permitting the evaluator to dispense with this method of adjustment based on the date of the offender’s last offense. Put simply, Dr. Tomich’s suggestion that Mr. Shields’ risk of reoffending should not be adjusted to account for his age is simply an arbitrary methodological choice without any empirical or other support.

Finally, the language Dr. Tomich uses in his report mischaracterizes the data regarding age and recidivism. For example, Dr. Tomich asserted that “child molesters are known to reoffend well into their fifties.” A more accurate, and significantly more modest, statement is that after age 60 the risk of reoffending by child molesters is extremely small. Hanson’s 2006 age study states that “in all the analysis (with or without controlling for Static-99 risk factors), there was a steady

decline in recidivism rates for offenders after the age of 40 years. The five-year recidivism rate of offenders over 60 was only 2%, compared with 14.8% for offenders less than 40.” Exhibit G, *supra* at 351.

B. The Invalidated Factors of Sex Offender Treatment, Substance Abuse, and Victim Status

After grounding his analysis in the “data-point” of the Static-99, Dr. Tomich then considered other dynamic factors. As discussed below, all of these factors have been demonstrated to bear no relationship to the risk of sex offender recidivism.

i. “Refusal” of Sex Offender Treatment

In his report Dr. Tomich stressed Mr. Shields’ failure to engage in sex offender treatment while incarcerated. Dr. Tomich noted that, although Mr. Shields was in treatment for depression with Dr. Graney, he “avoided” sex-offender specific treatment. Tomich Rep. at 10.¹⁸ Although Mr. Shields’ treating psychologist saw a “‘dramatic improvement’ in Mr. Shields’ ‘mood, functioning, self-view, interactions with others, insight, and motivation’”, Dr. Tomich declined to consider Mr. Shields’ participation in psychotherapy as a protective factor because the treatment was not sex-offender specific. Tomich Rep. at 10. Dr. Tomich further opined that “symptoms of depression or Post-Traumatic Stress Disorder also do not lead one to perpetuate sexual crimes against children” Tomich Rep. at 14. Presumably, then, in Dr. Tomich’s opinion treatment for

¹⁸ The evidence that Mr. Shields “avoided” treatment consists of the fact that he did not take advantage of it when offered, stating “I am dealing with a lot of mental health issues and do not feel that I could (sic) my full attention and focus on the program.” His clinician, Dr. Graney, agreed with this decision. Tomich Rep. at 10.

those disorders would be ineffectual in reducing the risk of recidivism.¹⁹ Dr. Tomich further opined that pedophilia is a “chronic condition” that “does not abate over time.” Tr. at 47. He therefore concluded that “[w]ere Mr. Shields to leave the Bureau of Prisons at this time he would be leaving as an untreated sexual offender who presents with a history of engaging in sexual crimes against children.” Tomich Rep. at 15.

Tomich’s emphasis on Mr. Shields’ lack of sex-offender treatment effectively confuses a risk-reducing factor with a risk-increasing factor. Although studies of offenders who completed voluntary outpatient treatment programs did show some reduction in recidivism risk, the reverse was not true - there was no empirically demonstrated increase in risk for individuals who did not participate in treatment. See Hanson, et. al., *Evaluating Community Sex Offender Treatment Programs: A 12-Year Follow Up of 724 Offenders*, 36 Canadian J. of Behavioral Sci., 87 (2004), introduced as Defendant’s Exhibit 7. See also Kriegman, Tr. at 229. In addition, there was no empirically demonstrated increase in risk for individuals who refused treatment as compared with individuals for whom treatment was simply not available. Tr. at 230. Simply put, treatment can help reduce the risk of recidivism, but the lack of treatment does not increase the risk of recidivism.

This supports Dr. Plaud’s opinion that “some professionals might regard Mr. Shields’ lack of participation in sexual offender specific treatment as indicative of a greater or significant likelihood that he might reoffend if released from a secure facility. While this may be colloquial or seemingly conventional wisdom, when the current empirical data and research on this very

¹⁹ Dr. Kriegman testified that there is no data to support Dr. Tomich’s conclusion that sex-offender specific treatment is more effective than any other type of treatment in reducing recidivism risk. Tr. at 236-237.

issue are understood, it is very clear that the data do not support the premise that refusal to participate in treatment has a significant relationship with sexual offense recidivism.” Plaud Rep. at 16-17.

Dr. Tomich’s emphasis on lack of sex offender treatment is typical of his approach to the risk assessment process and demonstrates that, although he labels his methodology “adjusted actuarial,” he in fact simply uses factors that he believes are clinically significant and disregards the total lack of empirical validation for his assumptions.

To further illustrate, the Court discussed with Dr. Tomich the “markers” that he uses to assess continuing risk in men who have been committed. The list he provided makes it clear that he is using a clinical method, not the adjusted actuarial method or the guided/structured clinical method, because none of the markers that he relies on has been empirically validated:

Well, a marker for me, if I can present a hypothetical situation, would be somebody that may have been committed as a sexually dangerous person, engaged fully in the treatment program, provided records that would indicate that that individual understood the nature of his sexual offending, understood why he had engaged in the sexual offending, had developed some kind of relapse prevention plan to forestall future offending, developed an adequate release plan that would provide for that individual’s and other’s safety in the community. So I think I’d require that as a minimum.

I’d also like to understand the individual on a more personal level. I’d like to interview the individual and have them explain to me what led them to be civilly committed in the first place, and what leads them now to believe that they no longer require civil commitment today.

Tr. at pp. 45.

Thus, the markers that Dr. Tomich uses to determine the existence of risk are: 1) offender engaged fully in the treatment program; 2) offender understands the nature of his sexual

offending; 3) offender developed a relapse prevention plan; and 4) Dr. Tomich's understanding of the offender based on a clinical interview. The first three factors have not been shown to have any relationship with recidivism. The last factor – the clinical interview -- has actually been shown to be *negatively* correlated with accurate predictions of future sexual dangerousness.²⁰

ii. Substance Abuse, Depression, and Victim Status

In his report Dr. Tomich also stressed Mr. Shields' history of alcohol and drug abuse, depression, and his status as a victim of sexual assault. "This evaluator also examined dynamic factors that might play a role in Mr. Shields' risk analysis. Mr. Shields presents with the dynamic factor of having a history of drug and alcohol abuse, depression, and a history of being a victim of sexual abuse and is dynamic in that it appears to be part and parcel of his depressive presentation. While it may be the case that his drug and alcohol abuse disinhibited him at the time of his offending, drug and alcohol abuse do not lead individuals to perpetuate sexual assaults on minors. Similarly, symptoms of depression or Posttraumatic Stress Disorder also do not lead one to perpetuate sexual crimes against children." Tomich Rep. at 14.

Again, none of these factors is correlated with sex offender recidivism. In 1998, Karl Hanson published a "meta-analysis" of sixty-one sex offender studies to identify the factors most strongly correlated with recidivism. See R. Karl Hanson & Monique Bussiere, *Predicting*

²⁰ Dr. Kriegman's affidavit details the results of a recent meta-analysis of 136 studies. This meta-analysis found a negative correlation with accurate predictions when the evaluator had access to a clinical interview. See Kriegman Aff., Exhibit 2 (Clinical Predictions) quoting William Grove, et. al., *Clinical versus Mechanical Prediction: A Meta-Analysis*, 12 Psychological Assessment at 19 ("The only design variable [of the studies] that substantially influenced the relative efficacy of the mechanical - and clinical - prediction methods was whether the clinicians had access to a clinical interview. *Alas, clinical predictions were outperformed by a substantially greater margin when such data was available to the clinician.*")

Relapse: A Meta-Analysis of Sexual Offender Recidivism Studies, 66 J. Consulting & Clin. Psy. 348 (1998), introduced as Defendant's Exhibit 8. This meta-analysis showed that alcohol abuse, substance abuse, depression, and victim status had no or almost no correlation with recidivism.²¹ This accords with Dr. Kriegman testimony that "...one of the surprising findings of Hanson's first meta-analysis, is that the correlation between a history of drug abuse or alcoholism and sex offense recidivism is zero, almost a perfect zero." Thus, these factors are invalidated and should never be used in risk assessment. Tr. at 245; 247.

4. Dr. Tomich's Methodology Amounts to Unstructured Clinical Judgment.

Dr. Tomich's methodology amounts to unstructured clinical judgment. He declines to adjust the Static-99 to account for the only empirically validated dynamic factor. The other "data-points" he uses -- lack of sexual offender treatment, depression, history of alcohol abuse, and victim status -- have been shown to have no correlation with recidivism. See footnote 21, *supra*.²²

In contrast to Dr. Tomich's methodology, Dr. Kriegman explained how an evaluator who is attempting a clinical variation of risk assessment should go about the evaluation:

²¹ Page 353 of the study provides a table listing the factors that were studied while page 351 explains how to read the table. The variable "r+" is used to designate the correlation with sexual offender recidivism. "Correlation rates greater than .30 would be considered large (recidivism rate differences of 30%), correlations greater than .20 moderate, and correlations in the .10 to .20 range small. Correlations less than .10 would have little practical utility in most settings." *Id.* at 351. A correlation rate of .00 is equal to chance. Thus, the table on page 353 indicates a correlation rate for depression of -.01, for victim status of -.01, and substance abuse of .03. These are correlations that are almost equal to chance. The correlation rate for alcohol abuse was .00, which is exactly equal to chance. Exhibit 8, at 351, 353.

²² Dr. Tomich admitted that he used factors that were not validated, although he asserted that "some" of his data points were validated. Tr. at 180. He did not, however, specify which were validated or provide any empirical support for this assertion.

Then there's the guided clinical or empirically guided clinical method, in which you don't use your clinical judgement as to what to consider. You take a list of factors that have been empirically validated as having real correlation with recidivism. You define them exactly as they have been described in the literature, that then validated their and showed that it had some real relationship to recidivism. So you make a list of them that you're going to consider before you do an evaluation. You may have 20 factors that have been found to be valid and truly related to recidivism, and you use the definition in the research studies. And you use the same list every time. You don't cherry-pick to figure out which factors you're going to use, this case or that case. And then you're free in the guided clinical or empirically guided approach to weight them as you choose using your clinical judgment. And as we heard testimony, and I believe its true, that has shown a validity about halfway between pure clinical judgment, which is virtually useless but has slight validity, and the actuarial approach.

Tr. at pg. 205-06.

Dr. Tomich did not follow this method. He did not begin with a fixed list of factors, he did not define factors in advance, and he did not have a method for weighing or combining those factors into an assessment of risk. In fact, Dr. Tomich was unwilling or unable to describe how he weighs various risk factors against one another:

Q: How do you tell what weight to give the particular other data points you claim to use?

A: I don't. I leave that in the discussion section and I leave the weight to give to the fact finder.

Tr. at 180.

Q: What is the relative weight you give to your actuarial risk assessment vis-a-vis all the other data points you claim are relevant to your risk assessment?

A: I give it significant weight. I think, unless there's reasons to adjust the actuarial score lower or higher, it should be given all the weight it accords . . ."

Tr. at 182-183.

Dr. Tomich's method essentially consists of reporting the score of the Static-99 and his "clinical judgment" about Mr. Shields' history - clinical judgment that, as described above, often accords significant weight to factors that have little or no correlation with sex offender recidivism. "I use the Static-99 and I use that as a specific data point, and I will, if I'm asked, let the jury know what the score is, whether that's high, moderate, low risk. But I'll also say that there are these other factors that I consider, dynamic factors, protective factors, historical factors from the record, his current understanding of his sexual offenses, whether he's had treatment or not. All those things play a part in my forming an opinion and I try to put those all forth in my discussion section of my report. And then the fact finder can either agree with me or not." Tr. at 84.²³

As Hanson has described, this is an "unstructured clinical judgment" approach. "Evaluators using the unstructured clinical approach integrate diverse material based on theory and their experience with similar cases. In such evaluations, neither the risk factors considered nor the method of combining the risk factors are fixed, and are allowed to change from case-to-case." Hanson, *Sexual Offender Recidivism Risk*, Exhibit B, *supra* at 158. In 2007, Hanson noted that, although different evaluators may have different approaches to unstructured clinical judgment, the commonalities defining the method are that "the risk factors are not specified in advance, nor is the method of combining the risk factors into an overall evaluation of risk." Hanson, *Meta-Analysis 2007-01*, Defendant's Exhibit 5, *supra*, at 3.

²³ Dr. Tomich's testimony also made it clear that he believes that the clinical method is a valid approach for predicting future dangerousness "in some cases." Tr. at 173. However, when pressed by the Court he was unable to cite any peer-reviewed literature finding that clinical judgment was a reliable methodology, and conceded that it was not the standard in the field. *Id.*

In 2007, Hanson again found that the accuracy of unguided professional judgment was consistently low, and concluded that, “unstructured professional judgment has been the traditional loser in the actuarial-clinical debates.” Id. at 3, 14.

IV. THE ADMISSION OF EXPERT TESTIMONY PREDICTING FUTURE SEXUAL DANGEROUSNESS WOULD VIOLATE DUE PROCESS.

As argued in detail, *supra*, expert testimony regarding risk assessment of sex offenders is insufficiently reliable to be admitted under Daubert. Daubert, of course, requires only that the proponent of such evidence establish its relevance and reliability by a preponderance of the evidence before a court may admit it. See Daubert, 509 U.S. at 593 n. 10 (citing Bourjaily v. United States, 483 U.S. 171, 175-76 (1987)); Cook ex rel. Estate of Tessier v. Sheriff of Monroe County, Fla., 402 F.3d 1092, 1107 (11th Cir. 2005); Bourne v. Town of Madison, 2007 WL 1447672 at 3 (D.N.H. 2007). Section 4248, however, mandates that the government must prove that Mr. Shields is a sexually dangerous person by *clear and convincing* evidence – a significantly more stringent standard than a preponderance burden. If, therefore, expert testimony bearing on the issue of sexual dangerousness cannot be shown to be reliable even under the lesser preponderance standard, then *ipso facto* it can never reliably prove sexual dangerousness by the more rigorous clear and convincing evidence standard of section 4248.

Since evidence of scientific risk assessment cannot demonstrate sexual dangerousness by the statutorily required clear and convincing evidence standard, the admission of expert testimony on that subject violates due process. Simply put, the admission of expert testimony known to be

inadequate to meet the burden of proof in this proceeding will violate due process because it will subject Mr. Shields to an unjustifiable risk that he will be erroneously deprived of his liberty.

Barefoot's holding that the admission of psychiatric opinions of future dangerousness do not violate due process is not controlling here; Barefoot's facts are materially distinguishable. In Barefoot, the psychiatrist only rendered an opinion on whether the defendant was likely to commit further criminal acts of violence, thereby posing a continuing threat to society. Barefoot, 463 U.S. at 894. Similarly, in Jurek v. Texas, which approved jury consideration of future dangerousness as an aggravating factor in capital sentencing, the issue was strictly confined to prognostications of future dangerousness, and no psychiatric testimony was offered. See Jurek, 428 U.S. 262, 267 (1996). Thus, neither the psychiatrist in Barefoot nor the jury in Jurek were required to diagnose a mental illness or abnormality in the defendant, and then draw a causal connection between the diagnosis and the probability that the defendant would be unable to stop himself from committing criminally violent acts.

Section 4248, by contrast, expressly requires proof by clear and convincing evidence that Mr. Shields suffers from a mental illness or abnormality, *and* that the mental illness or abnormality will *cause* him to have serious difficulty from refraining from child molestation if he is released. See 18 U.S.C. § 4247(a)(6). These statutory requirements of: 1) a psychiatric diagnosis; and 2) a causal connection between the diagnosis and sexual dangerousness, distinguish the issue in this case from the issues presented to the Supreme Court in Barefoot and Jurek.

Indeed, subsequent to Barefoot the Supreme Court has dealt with the intersection of psychiatric diagnosis and prognostication with legal burdens of proof. The Court implicitly

recognized the distinction between the issues posed by pure predictions of future dangerousness, as in Barefoot and Jurek, and the quite different issues posed by psychiatric predictions of future dangerousness that are alleged to be causally linked to a mental illness, which is the issue under section 4248. In Heller v. Doe the Supreme Court held that Kentucky could use a clear and convincing burden of proof in commitments of the mentally retarded, while requiring proof beyond a reasonable doubt as a predicate to the commitment of the mentally ill. See Heller, 509 U.S. 312, 321-22 (1993). The Heller Court upheld the difference in burdens of proof because the uncertainty and frequent inaccuracy of psychiatric diagnoses and predictions of future dangerousness, as compared with the relatively easy diagnosis of mental retardation, warranted the imposition of an elevated burden of proof to guard against erroneous psychiatric commitments. See Id. at 324.

In the years since Heller, several courts have employed reasoning similar to the Heller Court's to exclude supposedly scientific testimony where the testimony lacked relevance and reliability and threatened to produce an erroneous result.

In United States v. White Horse, for example, the court affirmed the exclusion by the district court of a psychologist's testimony that a Native American defendant accused of sexually abusing his young son was not sexually interested in boys. The psychologist's testimony rested, in part, on an "Abel Assessment," a purportedly scientific test involving the viewing of images of clothed and partially clothed people and the completion of a questionnaire designed to probe sexual behavior. Using these techniques, the "Abel Assessment" supposedly permitted the psychologist to predict the likelihood that the defendant was sexually interested in children under fourteen, and whether he would dissemble about his interest. White Horse, 316 F.3d 769, 774-75 (10th Cir.2003). The court of appeals agreed with the district court that the Abel test was not reliable nor

a good fit for the White Horse case because its use with Native Americans had not been properly studied, the photos shown to the defendant did not include images of Native Americans, and a study had shown that two of the Abel Assessment's three predictive equations had not been tested with incest offenders. Id.; Cf. also United States v. Velarde, 214 F.3d 1204 (10th Cir. 2000) (district court's failure to determine reliability of pediatrician's opinion that alleged child abuse victim's behavior and statements were consistent with an honest account of abuse required new trial); Doe ex rel. Rudy-Glanzer v. Glanzer, 232 F.3d 1258, 1266 (9th Cir. 2000) (penile plethysmograph test not reliable enough to be admitted); United States v. Powers, 59 F.3d 1460 (4th Cir. 1995) (penile plethysmograph not reliable or relevant enough to be admitted in incest prosecution).

In this case, the evidence cited, *supra* and in the exhibits and testimony that have been submitted to the Court, demonstrates that actuarial predictions of sexual dangerousness will be wrong approximately 66% of the time. The evidence further shows that actuarial predictions, even with their strikingly high rate of error, are by far the *most* accurate method of predicting sexual dangerousness. Thus, any expert opinion that Mr. Shields is sexually dangerousness is vastly more likely to be wrong than it is to be correct. Evidence having an error rate of well over 50% simply cannot be termed "clear and convincing," and it surely cannot justify the "massive curtailment of liberty" called for by section 4248. Vitek v. Jones, 445 U.S. 480, 491 (1980) (internal citation omitted).

V. "HIGH RISK" AND SIMILAR TERMS LACK RELEVANCE AND PROBATIVE VALUE, AND ARE UNFAIRLY PREJUDICIAL, CONFUSING, AND MISLEADING.

Toward the conclusion of the Daubert hearing, the Court expressed concern about the use of terms like “high risk” to describe Mr. Shields’ score on the Static-99, asking “...how do I have it presented to a jury in a way that’s not unduly prejudicial; in other words, using high risk instead of the 52 percent, or something that at least makes it understandable and not unduly prejudicial?”

Assuming the Court decides to admit the Static-99, its concern is well-founded. “High risk” and like terms lack probative value as descriptors either of Mr. Shields’ score on the Static-99, or of his risk of recidivism generally, and their use will mislead, confuse, and prejudice the jury.

Rule 403 of the Federal Rules of Evidence provides, in pertinent part, that “[a]lthough relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury...” Fed. R. Evid. R. 403. Under Rule 403, the Court must engage in a balancing test, weighing any probative value of the objected to evidence against the potential it may have to inflame the emotions of the jury, obscure the issues in the case, or mislead the jury. See Galarneau v. Merrill Lynch, Pierce, Fenner, and Smith, Inc., 504 F.3d 189, 205-06 (1st Cir. 2007). Should the Court determine that “the untoward effects of the proffered evidence” outweigh its probative value, it must be excluded. Faigan v. Kelly, 184 F.3d 67, 80 (1st Cir. 1999).

Concerning the use of terms like “high risk,” the following colloquy occurred between the Court and Dr. Kriegman:

THE COURT: But would the most accurate way to discuss Static-99, if you were to present it to a jury or to me, to simply, instead of saying high risk, describe it the way that it was verbally described, which is a score of

6, people with scores of 6, 52 percent of them? Is that an accurate way of describing it?

THE WITNESS: That would be far, far, far more accurate than using words like “high.” Carl Hanson did not create those labels with an eye towards sexual dangerousness statutes. He was just saying, let’s take this data and let’s divide this group, this group, we’ll call them middle group or high group, and then let’s go see what they actually do.

Tr. at 222.

Thus, the evidence before this Court is that the designer of the Static-99 did not create the labels associated with that instrument – e.g., “high risk” and “moderate high risk” – for use in civil commitment proceedings. Instead, he made a utilitarian decision to divide the subjects of his study into groups according to their score on the Static-99, and then chose labels to assist in easily tracking the conduct of each group. The labels are arbitrary in the sense that their only “real” referent is a subject’s membership in a group having a particular score on the Static-99; they do *not* denote any inherent psychological or character trait of any specific subject. To use them as though they do describe the character or condition of a specific individual, as the government proposes to do, is an abuse of Hanson’s terminology.

And, because the labels do not describe anything other than a methodological convenience, they are utterly without relevance to this case. Use of the labels simply has no tendency to make any fact that is of consequence to the determination of this action more or less likely, and their admission would therefore violate the most basic rule of evidence. See Fed. R. Evid. 401; Cf. United States v. Grande, 353 F.Supp.2d 623, 640 (E.D. Va. 2005) (in capital case, striking language alleging defendant “continued to conduct and influence...gang business...” from government’s

supporting sentencing information because language so vague and undefined that it lacked probative value on the issue of future dangerousness and would prejudice, confuse, and mislead the jury). Even if, however, the Court finds that the Static-99 labels have some probative value, that probative value is far outweighed by their potential to unfairly prejudice, confuse, and mislead the jury.

“Unfairly prejudicial evidence...is evidence that triggers the mainsprings of human action in such a way as to cause a jury to base its decision on something other than the established proposition in the case.” United States v. Currier, 836 F.2d 11, 18 (1st Cir. 1987) (internal citation omitted). It is difficult to imagine that, in the social, political, and media climate that has spawned legislation like section 4248, there is a more prejudicial epithet with which Mr. Shields could be saddled than “high risk” sex offender. And, inasmuch as the term “high risk” does not refer to anything essential about Mr. Shields, but merely describes his similarity, in limited respects, to an arbitrarily labeled group of sex offenders, the epithet is totally inaccurate.

For the same reason, the epithet will also confuse and mislead the jury. The relevant issue in this case is whether it can be proved, by clear and convincing evidence, that Mr. Shields suffers from a mental abnormality that will make it seriously difficult for him to refrain from child molestation if he is not incarcerated. Calling Mr. Shields “high risk” obscures that issue by distracting the jury’s attention from what the Static-99 *actually says* – that Mr. Shields received a certain score, and that a group of offenders similar to Mr. Shields in limited, narrowly defined respects reoffended at a particular rate during a particular period of time subsequent to their release from incarceration. To suggest, as the government proposes to do, that Mr. Shields is a “high risk” offender is simply to *misstate* the meaning of his score on the Static-99, and thereby to confuse and mislead the jury.

VI. EXPERT TESTIMONY REGARDING FUTURE SEXUAL DANGEROUSNESS WILL INORDINATELY INFLUENCE THE ADVISORY JURY.

Study after study has shown that clinical judgment is no more accurate at predicting future dangerousness than a juror's lay opinion. And, as described *supra* and in the various exhibits that have been submitted to this Court, even methods of risk assessment that produce more reliable predictions than clinical judgment are "reliable" only in the most limited sense of that word.

Nevertheless, jurors, faced with the responsibility of determining whether an individual who has already committed at least one act of child molestation will do so again, and undoubtedly concerned with the consequences of an incorrect decision, are understandably likely to defer to an "expert" determination confirming their cautious inclinations, even if its reliability is questioned by another "expert." Given the choice between an expert who says that an individual would have serious difficulty refraining from child molestation if released, and one who merely says that no such prediction can be made, members of the jury surely will be tempted to opt for the expert who claims he can help them in performing their duty, and whose opinion suggests dire consequences if the defendant is not civilly committed.

As Judge Garza noted in Flores v. Johnson, there is good reason to fear that the testimony of an "expert" on the issue of predicting future sexual dangerousness will be given more weight than it deserves:

The problem here (as with all expert testimony) is not the introduction of one man's opinion on another's future dangerousness, but the fact that the opinion is introduced by one whose title and education (not to mention designation of "expert") gives him significant credibility in the eyes of the jury as one whose opinion comes with the imprimatur of scientific fact. As has been previously recognized, when a medical doctor testifies that future dangerousness is a scientific inquiry on which they have particular expertise, and testifies that a particular

defendant would be a “continuing threat to society” juries are almost always persuaded.”

Flores v. Johnson, 210 F.3d at 466. See also APA Amicus Brief Barefoot v. Estelle, Defendant’s Exhibit 4, *supra*, (“A psychiatrist comes into the courtroom wearing a mantle of expertise than inevitably enhances the credibility, and therefore the impact, of the testimony . . . when a prediction of future dangerousness ‘is proffered by a witness bearing the title ‘doctor,’ it’s impact on the jury is much greater than if it were not masquerading as something it’s not.”) (internal citations omitted).

Dr. Tomich’s opinion amounts to unstructured clinical judgment “masquerading” as “scientific, technical, or other specialized knowledge.” Fed. R. Evid. 702. Any expert opinion regarding sex offender risk assessment will be little more “scientific” and reliable than Dr. Tomich’s discredited approach. Hence, because all such testimony has the very real potential to inordinately influence the jury, it must be excluded.

JEFFREY SHIELDS

By his attorney,

/s/ Page Kelley

Page Kelley

B.B.O. #548237

Federal Defender Office

408 Atlantic Avenue, 3rd Floor

Boston, MA 02110

Tel: 617-223-8061

CERTIFICATE OF SERVICE

I hereby certify that this document filed through the ECF system will be sent electronically to the registered participants as identified on the Notice of Electronic Filing (NEF) and paper copies will be sent to those indicated as non registered participants on January 11, 2008.

/s/ Page Kelley

Page Kelley